

<https://helda.helsinki.fi>

Bio CRM: A Data Model for Representing Biographical Data for Prosopographical Research

Tuominen, Jouni Antero

CEUR Workshop Proceedings
2018

Tuominen , J A , Hyvönen , E A & Leskinen , P 2018 , Bio CRM: A Data Model for Representing Biographical Data for Prosopographical Research . in Proceedings of the Second Conference on Biographical Data in a Digital World 2017 (BD2017) . CEUR Workshop proceedings , vol. 2119 , CEUR Workshop Proceedings , pp. 59-66 , Conference on Biographical Data in a Digital World , Linz , Austria , 06/11/2017 . < <http://ceur-ws.org/Vol-2119/paper10.pdf> >

<http://hdl.handle.net/10138/312790>

unspecified
publishedVersion

Downloaded from Helda, University of Helsinki institutional repository.

This is an electronic reprint of the original article.

This reprint may differ from the original in pagination and typographic detail.

Please cite the original version.

Bio CRM: A Data Model for Representing Biographical Data for Prosopographical Research

Jouni Tuominen^{1,2}, Eero Hyvönen^{1,2}, and Petri Leskinen¹

¹Semantic Computing Research Group (SeCo), Aalto University, Finland and

²HELDIG – Helsinki Centre for Digital Humanities, University of Helsinki, Finland

<http://seco.cs.aalto.fi>, <http://heldig.fi>

firstname.lastname@aalto.fi

Abstract

Biographies make a promising application case of Linked Data: they can be used, e.g., as a basis for Digital Humanities research in prosopography and as a key data and linking resource in semantic Cultural Heritage (CH) portals. In both use cases, a semantic data model for harmonizing and interlinking heterogeneous data from different sources is needed. This paper presents such a data model, Bio CRM, with the following key ideas: 1) The model is a domain specific extension of CIDOC CRM, making it applicable to not only biographical data but to other CH data, too. 2) The model makes a distinction between enduring unary roles of actors, their enduring binary relationships, and perduring events, where the participants can take different roles modeled as a role concept hierarchy. 3) The model can be used as a basis for semantic data validation and enrichment by reasoning. 4) The enriched data conforming to Bio CRM is targeted to be used by SPARQL queries in a flexible ways using a hierarchy of roles in which participants can be involved in events.

Keywords: Linked Data, Data models, Biographical representation, Event-based modeling, Role-centric modeling, Prosopography

1. Event-based Approach for Biographies

The underlying idea of this paper is to represent life stories of people as Linked Data, extracted and aggregated from heterogeneous distributed data sources, such as dictionaries of national biographies, museum collections, library databases, Wikipedia etc. (Hyvönen et al., 2018). Linked biographical data facilitates studying enriched individual life stories in biography research (Roberts, 2002) as well as in prosopography research on groups of people (Verboven et al., 2007; Keats-Rohan, 2007). This paper addresses the fundamental technical research question that has to be solved in this kind of work: *how to model life stories, so that they can be enriched from heterogeneous data sources and interlinked with each other in a semantically interoperable way?*

Our research hypothesis is that a good choice for data modeling and harmonization is the event-based approach where a person's life is seen as a sequence of spatiotemporal, interlinked events from birth to death—a person may also be involved in prenatal and posthumous events. For example, metadata about a painting in a gallery actually means that there has been a painting event, and this could be included in the timeline of the artist's semantic biography. Event-based modeling and ontologies have already been found useful for harmonizing heterogeneous cultural heritage data. A most notable and widely used ontology for this is the CIDOC Conceptual Reference Model (CRM)¹ (Doerr, 2009), but there are also other models (Raimond and Abdallah, 2007; Scherp et al., 2009; Shaw, 2010; van Hage et al., 2011).

A recurring problem in event based modeling is, however, that it is not necessarily clear what is an event, since many

relations and roles (Kozaki et al., 2006) occur in time and space, too. For example: Are family relationships events, e.g., being the father of or being married to someone? Are professions events, such as being a president of a country, because holding an office occurs in time and space with an agent involved?

For example, the concept of “bishop” would be useful in representing and querying biographical data, but what does being a bishop actually mean? Is there a class and instances of bishops, is being a bishop a property of a person or a role, or how does the concept relate to the event of holding a bishop's office? Obviously, being a bishop can be represented in different ways, but then harmonizing and querying of data about bishops becomes very difficult since the user cannot be sure in what alternative ways being a bishop is actually represented. On the other hand, we clearly need foundational ontological structures (Guarino and Welty, 2002) for representing pieces of heterogeneous knowledge in a systematic and unique way, but on the other hand, there is a need for simple conceptualizations and property structures for querying the data and representing the data for the human user.

To address these problems, this paper introduces a data model, Bio CRM, for harmonizing, enriching, and using biographical linked data based on events. Bio CRM is an extension of CIDOC CRM to the biographical domain. This ISO standard was selected as the basis because it is the most widely used ontology standard for event-based modeling in museums and has been integrated with the Functional Requirements for Bibliographic Record (FRBR) family of modeling standards in libraries². Data from museums and libraries are essential in describing life stories.

¹<http://cidoc-crm.org>

²<https://www.ifla.org/about-the-frbr-review-group>

In the following, major use cases for Bio CRM in biographical and prosopographical research are first listed. After this, the design principles and the actual data model are presented, with three online applications illustrating the use of the system. Finally, related work is discussed with a comparison of Bio CRM with other data models.

2. Prosopographical Method

The aim of using the Bio CRM data model in our case studies is to facilitate using the prosopographical research method (Verboven et al., 2007, p. 47) that consists of two major steps. First, a target population of people is selected that share desired characteristics for solving the research question at hand. For example, our research question may be related to social networks of men who were born in Finland 1800–1900 and were artists. Second, the target group is analyzed further, and compared with other groups, in order to solve the research question.

1. Determine target groups Target groups can be found by data filtering with a human in the loop. For example, SPARQL SELECT can be used to create a tabular set of selected instances. In our case, faceted search is a promising option for filtering out target groups in a flexible and dynamic way. An interesting possibility for further research would be to try to do filtering automatically using knowledge discovery.

Once a target group has been determined, specific working hypotheses and specific historical questions concerning the group can be formulated and analyses performed.

2. Prosopographical analysis Linked data and SPARQL querying provides many possibilities for analyzing target group data. For example, it is possible to analyze the structure and changing composition of the group in time and the changing roles of individuals or subgroups. In this case the result of a SPARQL SELECT or CONSTRUCT is analyzed further by specific algorithms or visualization tools, such as information graphics.

Another option is to employ methods of network analysis methods and tools (Easley and Kleinberg, 2010; Hanneman and Riddle, 2005) and visualizations (Dadzie and Rowe, 2011; Kehrer and Hauser, 2013). In this case, for example, a SPARQL CONSTRUCT can be used for creating an RDF network based on the underlying data.

3. Representing Biographies as Linked Data

To aggregate, enrich, and link biographical data with related datasets the data must be made semantically interoperable, either by data alignments (using, e.g., Dublin Core and the dumb-down principle) or by data transformations into a harmonized form (Hyvönen, 2012). In our case we selected the data harmonization approach and the event-centric CIDOC CRM ISO standard as the ontological basis, since biographies are based on life events. CIDOC CRM provides a common and extensible semantic framework for representing cultural heritage information, operating as a "semantic glue" for integration, mediation, and interchange of heterogeneous datasets from, e.g., museums, libraries, and archives. In our work, biographies are modeled as collections of CIDOC CRM events, where each event is characterized by the 1) actors involved, 2) place, 3) time, and 4)

the event type. Bio CRM extends CIDOC CRM by introducing role-centric modeling. The reason for the extension is that while CIDOC CRM does include a mechanism for representing roles of participants in events, its encoding in RDF is complex and still in experimental phase (see Section 5 for further discussion).

Bio CRM provides the general data model for biographical datasets. The individual datasets concerning different cultures, time periods, or collected by different researchers may introduce extensions for defining additional event and role types. The Linked Data approach enables connecting the biographies to contextualizing information, such as the space and time of biographical events, related people, historical events, publications, and paintings.

The core design principles of the data model are:

- The model is a domain specific extension of CIDOC CRM, making it applicable to not only biographical data but to other Cultural Heritage (CH) data, too.
- The model makes a distinction between enduring unary roles of actors, their enduring binary relationships, and perdueing events, where the participants can take different roles modeled as a role concept hierarchy.
- The model can be used as a basis for semantic data validation and enrichment by reasoning.
- The enriched data conforming to Bio CRM is targeted to be used by SPARQL queries in a flexible ways using a hierarchy of roles in which participants can be involved in events.

Bio CRM makes a clear distinction between person's attributes, relations between people, and events in which people participate in different roles.

- *Attributes* are properties of a person that are assumed to characterize her independently of time and space. For example, place and time of birth can be modeled as attributes.
- *Relations* are established between people and are assumed to characterize them independently of time and space. For example, father-of is such a relation. Relations can, however, have time and space as qualifiers, e.g., student-of. For example, Ferdinand Bol (1616–1680) was a student of Rembrandt in 1630–1641, starting his own studio in 1641, but can be characterized as a student-of Rembrandt in general. His years in Rembrandt's studio as a student can be represented as an event (see below), if needed.
- *Events* take place in time and space and involve participants in different roles, expressing the ways in which persons participate in events. For example, an officiant may participate in a certain baptism event.

The core classes and properties of Bio CRM are presented in Figure 1. The namespace of the Bio CRM schema is <http://ldf.fi/schema/bioc/>, here used with the prefix `bioc`. The full specification of Bio CRM (class and

property listing) is available in the namespace URI. Similarly, the prefix `cidoc` is used for CIDOC CRM's namespace <http://www.cidoc-crm.org/cidoc-crm/>. A central focus in representing biographical data is representing people and their networks. A person is represented as an instance of `bioc:Person`, a subclass of `cidoc:E21_Person`. This instance-of relationship is persistent and never changes during the life of the person. In order to identify a person, the person is associated with core data: appellations, i.e., names and identifiers in other data repositories, birth time and place, and death time and place, using CIDOC CRM. Person's birth and death are represented as a Birth/Death event, which can be qualified with time and place. Birth can also incorporate information about the mother and father.

In addition to the core data, a person can also have other attributes, relationships, and participate in events. Having a role, say Teacher, may be temporary or something inherent characterizing a person as a whole in all times, even if it is possible also to specify when exactly the role was present (e.g., a professorship). Being a Teacher by education is different from saying that the person happened to participate in a particular teaching event, e.g., gave a lecture, in the role of a Teacher.

Genders, nationalities, and occupations of people are represented by relating a person to a unary role using the property `bioc:bearer_of`. Figure 2 depicts an example of John F. Kennedy in the role of President. The role (blank node, as there is no need to give a identity to it in this case) is attached to the a Person using the `bioc:has_occupation` relation (subproperty of `bioc:bearer_of`). While this expresses the gender, nationality, or occupation generally, it's also possible to qualify the roles in time and and space by attaching a contextualizing event, e.g., the employment of a person. This is useful, as people have different roles during their life that typically perdure a shorter period of time and may have other qualifiers, too. For example, John Kennedy had the role of President in the US in 1961–1963.

The unary roles of Bio CRM are divided into the following class hierarchy:

```
bioc:Unary_Role
  bioc:Gender
  bioc:Nationality
  bioc:Occupation
```

The role class hierarchy can be further extended in individual datasets, e.g., by listing the prevalent occupations in a certain cultural era.

The same role-based pattern is used for representing inherent relationships between people, such as family relations (mother, cousin, aunt, etc.) and social relations (studentOf, knows, etc.). Relationships are represented by relating an actor (a person or group) to another actor in a role by using one of the subproperties of the property `bioc:has_relation`. The role is attached to the another actor with the property `bioc:inheres_in` (inverse property of `bioc:bearer_of`). Figure 3 depicts an example of John F. Kennedy having a spouse Jacqueline Kennedy Onassis. Similarly to unary roles, relationships

can be qualified with temporal and spatial information by using an event to contextualize the role. A person may have been some point a Spouse, a Lawyer in a company, and a President of a country, possibly several times at different occasions. For example, John Kennedy was Spouse of Jacqueline Kennedy Onassis in 1953–1963.

The binary roles of Bio CRM are divided according the following class hierarchy:

```
bioc:Binary_Relationship_Role
  bioc:Person_Relationship_Role
    bioc:Family_Relationship_Role
    bioc:Social_Relationship_Role
  bioc:Intergroup_Relationship_Role
  bioc:Group_Relationship_Role
```

Similar as for unary roles, the binary role classification can be extended in individual datasets.

The individual events of biographies are represented as subclasses of `bioc:Event` that is a subclass of `cidoc:E5_Event` inheriting its properties. From a semantic viewpoint, events are described especially in terms of

- the time of the event
(`cidoc:P4_has_time-span`),
- place of the event (`cidoc:P7_took_place_at`),
- actors that participated in it
(`cidoc:P11_had_participant`),
- other resources involved
(`cidoc:P12_occurred_in_the_presence_of`).

Time and place properties refer directly to time spans and instances of places, respectively. The values for participating actors and other resources are instances of role classes. An actor role associates an actor with a role, making it possible for a person to participate in events in different roles that can also be qualified in terms of additional properties. Similarly to actors, physical objects and immaterial things can be involved in an event in specific roles.

Events can be used for qualifying a unary (e.g., an occupation) or a binary relation further, i.e., in such cases an instance of `bioc:Event` has to be created. As an example, Figure 4 represents the presidency of John F. Kennedy qualified with time and the country. Another example in Figure 5 depicts the marriage of John F. Kennedy and Jacqueline Kennedy Onassis qualified with time. The individual datasets may introduce their own classifications of event types and associated roles.

By using roles, it is possible to keep the number of properties smaller, because different properties for different roles are not needed. Instead, different role classes are used. Such a model is simpler to query using SPARQL and provides the user with a set of useful and natural hierarchy of role concepts.

Possible roles that can be attached to certain event types are specified using the OWL restriction `owl:AllValuesFrom` on property `cidoc:P11_had_participant`. This can be used for

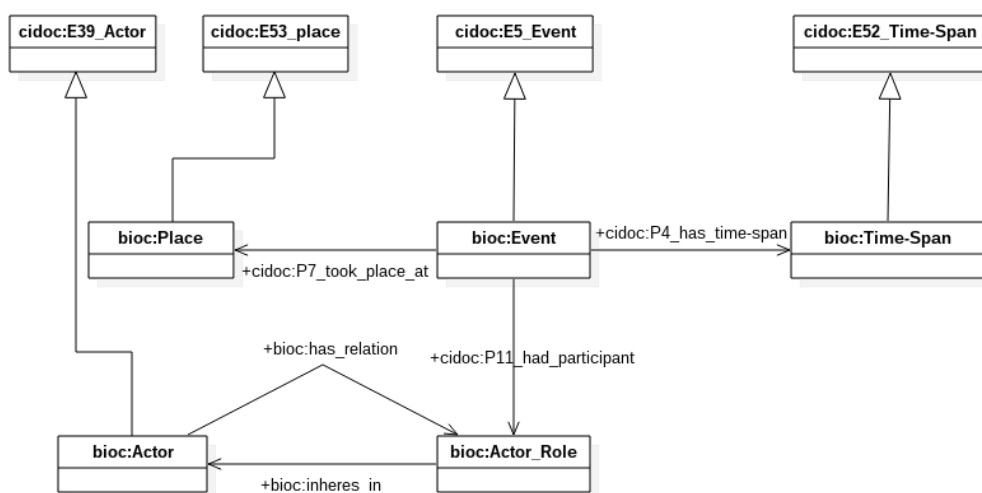


Figure 1: The core classes and properties of Bio CRM.

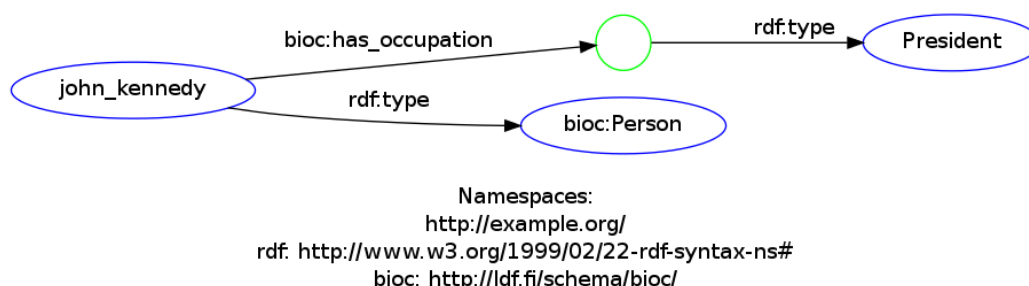


Figure 2: Unary role: John F. Kennedy as a president.

validating data, i.e., to see if the events have participants in incompatible roles. It is recommended that each event class, say *Baptism*, has a corresponding class of allowed roles, say *BaptismActorRole*. Its subclasses are roles whose instances can be used in filling the roles. In this way, the data annotator can be guided to use only the correct roles, and the new role class can be used for finding resources in roles when querying. The role hierarchy facilitates sharing roles between events and modifying their role structure easily by just editing the role hierarchy. This is more flexible than, e.g., changing property names, if roles were represented using different properties.

The following SPARQL query is an example for finding all "bishops" in a dataset. Note that because of the chosen role modeling approach, the query returns both the bishops as unary roles (occupation) and acting bishops in specific events (e.g., a confirmation). The namespace prefix declarations are omitted from the query for brevity.

```
SELECT ?person ?name ?event_title
      ?time ?place
WHERE {
  ?role a :Bishop ;
        bioc:inheres_in ?person .
  ?person a bioc:Person ;
          cidoc:P131_is_identified_by
            ?appellation .
  ?appellation rdfs:label ?name .
```

```
OPTIONAL {
  ?event cidoc:P11_had_participant ?role ;
         rdfs:label ?event_title ;
         cidoc:P4_has_time-span ?time ;
         cidoc:P7_took_place_at ?place .
}
```

4. Bio CRM Case Studies

In the following, three case studies for using Bio CRM are presented.

4.1. Early Modern Letters Online (EMLO)

Bio CRM was originally developed as a spin-off case study related to the database and web service Early Modern Letters Online (EMLO)³. EMLO is a collaboratively populated union catalogue of sixteenth-, seventeenth-, and eighteenth-century letters, created by the Cultures of Knowledge project⁴ at the University of Oxford. It brings manuscript, print, and electronic resources together in one space, increasing access to and awareness of them, and allows disparate and connected correspondences to be cross-searched, combined, analyzed, and visualized.

In addition to purely epistolary data, EMLO contains prosopographical information related to the people in

³<http://emlo.bodleian.ox.ac.uk>

⁴<http://www.culturesofknowledge.org>

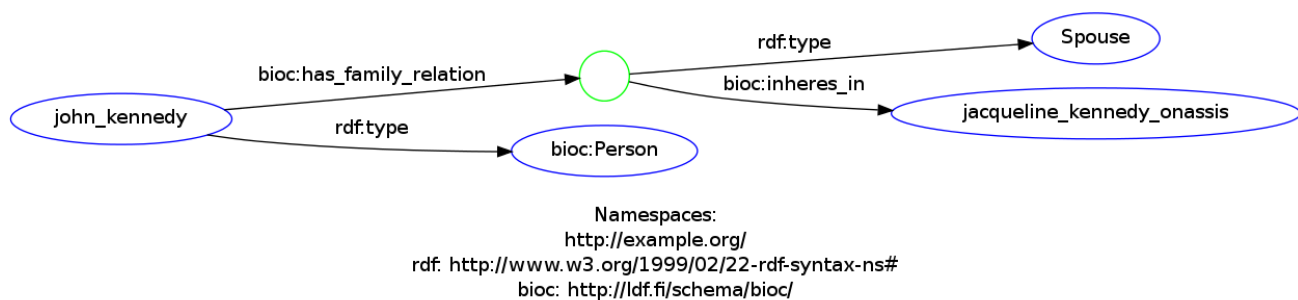


Figure 3: Relationship: John F. Kennedy has a spouse Jacqueline Kennedy Onassis.

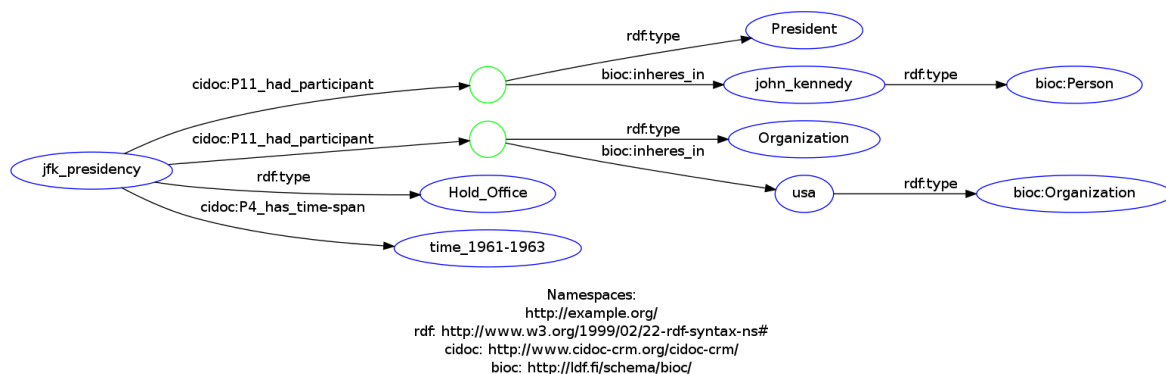


Figure 4: Event: qualifying the presidency of John F. Kennedy with time and the country.

the database, modeled as events and social relationships. Events cover activities that the people have participated in during their lives, such as birth and death, ecclesiastic and educational activities, creations of works, travels, and residences. The event metadata includes the event name, type, participants and their roles, time span, location, and source information. As a pilot Linked Data publication of the EMLO database (Tuominen et al., 2018), we have converted the prosopographical data into RDF format using CIDOC CRM for the event-based modeling and W3C's PROV model (Lebo et al., 2013) for representing the roles of participants in the events.

As a next step, we propose to convert the data into a Bio CRM representation, and build the event and role hierarchies pertaining to the activity types stored in the database. The top levels of the event hierarchy of the EMLO database are the following ones:

```
bioc:Event
  :Ecclesiastical_Event
  :Educational_Event
  :Political_Event
  :Professional_Event
  :Social_Status_Change
```

The class `:Ecclesiastical_Event` can be divided further into subclasses with attached roles, such as:

- `:Baptism` – `:Officiant`, `:Baptismal_Candidate`, `:Godfather`, `:Godmother`, `:Religion`

- `:Confirmation` – `:Officiant`, `:Confirmation_Candidate`, `:Religion`

4.2. The register of the high school "Norssi" alumni

Bio CRM has been applied in the study of transforming printed biographical registers into Linked Data and enriching their contents using Named Entity Linking. As a concrete case study, we have concentrated on the printed register "Norssit 1867–1992. Helsingin Norssin matrikkeli", a book of 708 pages, containing short bios of over 10 000 students and teachers of the prominent Finnish high school "Norssi", a training school of the University of Helsinki. The final application in use⁵ based on this case study is described in more detail in (Hyvönen et al., 2017), and the underlying data model is presented in (Leskinen et al., 2017).

4.3. Semantic National Biography of Finland

Bio CRM has also been used in creating the first prototype demonstrator of the Semantic National Biography of Finland (Hyvönen et al., 2018), based on a collection of some 13 000 short biographies that were transformed into RDF-format and enriched using data linking to external datasets. Application of Bio CRM to prosopographical research in the Norssi and Semantic National Biography case studies is described in more detail in (Leskinen et al., 2018).

⁵The application is available at <http://www.norssit.fi/semweb/>. Its linked open data is published at the Linked Data Finland service <http://ldf.fi> at the SPARQL endpoint <http://ldf.fi/norssit/sparql>.

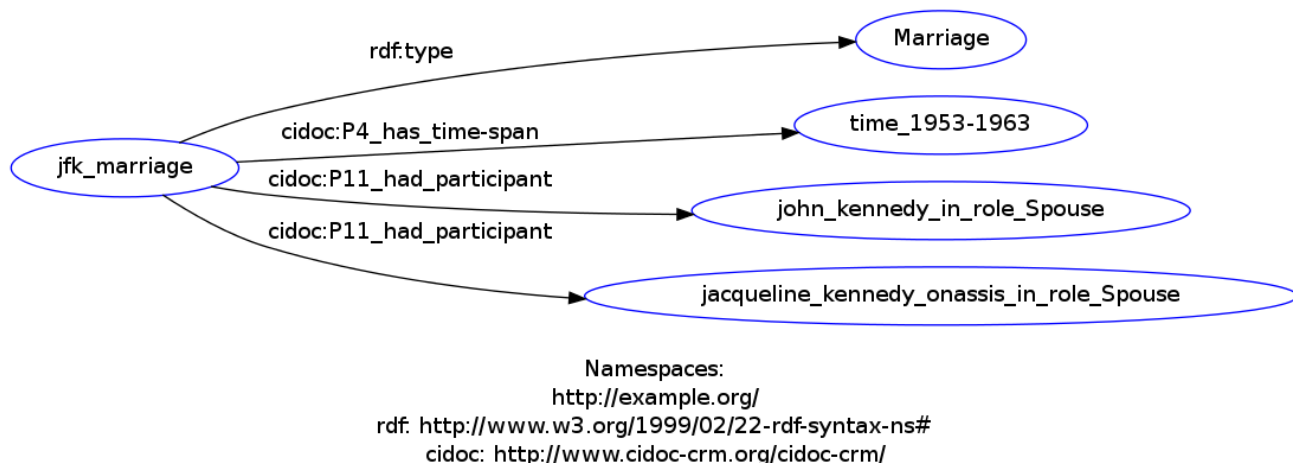


Figure 5: Event: qualifying the family relation between John F. Kennedy and Jacqueline Kennedy Onassis with time. Note that the roles are depicted in the figure as URI resources instead of blank nodes (see Figure 3) due to space constraints. For more detailed figure, see <https://goo.gl/Ex2Uu2>.

5. Discussion, Related Work, and Future Research

Bio CRM is to the best of our knowledge the first attempt to extend CIDOC CRM into the domain of biography and prosopography with additional subclasses, properties, and modeling guidelines. A major benefit of the model is the compatibility with cultural heritage data from museums, libraries, and archives represented using the same standard framework ontology.

From a modeling perspective, this paper presented the idea of making a distinction among attributes, relations, and events, where entities participate in different roles in a qualified manner, not as entities themselves. The underlying rationale for this is to harmonize the knowledge representation with fewer categories and at the same time keep the model expressive and easy to use by SPARQL queries. If needed, the model can be extended with transformational rules, by which more expressive and foundational event structures can be transformed into simpler attribute and role structures, when needed and appropriate from an application perspective, and vice versa. Events are needed for harmonizing data for the machine but the human end user often conceptualizes the world in document-centric or other ways. Thus, additional representations of the same event-based knowledge may be useful, especially if it is generated automatically by the system.

Our experiences of applying a first version of Bio CRM to the three case studies suggest the the model is usable for practical purposes. Though formal evaluation of the model has not been conducted, the application of Bio CRM to data originating from different sources, in different formats, and covering different eras is an indication of the suitability of the model to act as a general harmonizing model for prosopographical data. In the spirit of design science methodology, Bio CRM was designed to solve the modeling needs of the prosopographical data of the Early Modern Letters Online (EMLO), which provides a rich classification of the events and associated roles of people in the early modern

era. In the cases of high school "Norssi" alumni and Semantic National Biography, Bio CRM is used to model, e.g., the family relations and titles of the people (e.g., education, occupation). The idea of using semantically defined linked data for modeling and aggregating biographical and related data seems to be a promising approach for biography and prosopography. However, more work is needed in detailing out more precisely the class and property structures of the model in the general case—so far new classes and properties have been introduced based on the particular use cases and the general modeling principles presented in this paper.

Biographical data has been studied by genealogists (e.g., (Event) GEDCOM⁶), CH organizations (e.g., the Getty ULAN⁷), and semantic web researchers (e.g., BIO ontology⁸). Semantic web event models include, e.g., Event Ontology (Raimond and Abdallah, 2007), LODE ontology (Shaw, 2010), SEM (van Hage et al., 2011), and Event-Model-F⁹ (Scherp et al., 2009). Also, Bibliographic Ontology (BIBO) (D'Arcus and Giasson, 2009) includes a model for events. For a more detailed comparison on event models, see (Scherp and Mezaris, 2014). A history ontology with map visualizations is presented in (Nagypal et al., 2005), and an ontology of historical events in (Hyvönen et al., 2007). Visualization using historical timelines is discussed, e.g., in (Jensen, 2003), and event extraction reviewed in (Hogenboom et al., 2011).

PROSO (Zingoni, 2014) is a data model for presenting prosopographical data records. It has a strong focus on representing the provenance information of the records using factoids, and uses event-based modeling for stating the changes of a person (e.g., receiving a honorary ti-

⁶<http://en.wikipedia.org/wiki/GEDCOM>

⁷<http://www.getty.edu/research/tools/vocabularies/ulan/>

⁸<http://vocab.org/bio/>

⁹<https://www.kd.informatik.uni-kiel.de/en/research/ontologies/core-ontologies>

tle). Vocabularies and ontologies for representing personal relationships include the Standards for Networking Ancient Prosopographies (SNAP) project¹⁰ and the Relationship vocabulary (Davis, 2004). Existing data models that support role representation include CIDOC CRM, SEM, VIVO/BFO, Schema.org, PROV, and the Organization Ontology.

CIDOC CRM includes a mechanism for representing the role of an active participant in an event, modeling it as a property of the property that is used for representing the participating actor (see CIDOC's P14.1 in the role of). There is a proposal for encoding CIDOC's properties of properties as RDF¹¹, introducing new class for the property and auxiliary properties for connecting the event and the participant, which adds complexity to the data model. The status of the proposal is still experimental.

Simple Event Model (SEM) is a general model for expressing events, with support for three alternative representations for roles, based on using a) `rdf:value`, b) reification, or c) named graphs. All of the techniques add complexity to the data model. The model also introduces a new property `sem:roleType`, instead of using the standard RDF property `rdf:type`.

VIVO Integrated Semantic Framework ontology modules (VIVO-ISF)¹² include a model for representing roles in events. The model uses the properties of the Basic Formal Ontology (BFO) (Smith et al., 2015) and the related Relations Ontology (RO)¹³, `inheres in` to attach the role to a person, and `realized in` to attach the role to an event (and their inverse properties `bearer of` and `realizes`). It should be noted that the inclusion of the properties `inheres in/bearer of` and `realizes/is-realized in` in BFO is unclear in the current version (BFO 2.0)¹⁴. Bio CRM has taken some inspiration from VIVO, and uses similar properties `bioc:inheres in` and `bioc:bearer of`, but retains compatibility with CIDOC CRM by using the property `cidoc:P11.had participant` to attach the role to the event.

Schema.org¹⁵ provides a class `schema:Role` to represent additional information of a relation or property. It can be used to model the roles of participants in events. The instance of the `schema:Role` acts as an intermediary node between the event and participant, both of them attached with the "original" property, e.g., `cidoc:P11.had participant`. The model's strength is its simplicity, but the re-use of a property in such a way might be unintuitive. The model also introduces a new property `schema:additionalType`, instead of using the standard RDF property `rdf:type`.

¹⁰<https://snapdrgn.net>

¹¹<http://www.cidoc-crm.org/roles-in-the-cidoc%E2%80%9990crm-modelling-properties-of-properties>

¹²<https://wiki.duraspace.org/display/VIVODOC19x/Ontology+Reference>

¹³<https://github.com/oborel/obo-relations/>

¹⁴<https://github.com/BFO-ontology/BFO/blob/BFO2.0/README.md>

¹⁵<http://schema.org>

W3C's PROV model uses qualified associations to specify the roles of the participants in events. The instance of the class `prov:Association` is attached to the role using the property `prov:hadRole`, to the agent using the property `prov:agent`, and to the event using the property `prov:qualifiedAssociation`. Thus, the standard CIDOC CRM event can be qualified with such an association, but it might be unintuitive for the user to represent the qualifier separately from the `cidoc:P11.had participant` relation. By design, PROV is meant for representing provenance information involved in producing a piece of data or thing; all biographical events are not such activities.

W3C's Organization Ontology (Reynolds, 2014) is a core ontology for describing organizational structures. It includes a model for specifying membership roles people have in organizations by introducing the class `org:Membership`. Such a membership is associated with the role using the property `org:role`, to the agent with the property `org:member`, and to the organization with the property `org:organization`. The modeling approach is similar to the W3C's PROV model, but using different property names.

Bio CRM's property `bioc:inheres in` is used for representing both atemporal (unary roles and binary relationships without qualifiers) and temporal (qualified by using events) roles of people. This is an informed decision for the simplicity of the model. A different approach has been chosen in Basic Formal Ontology (BFO) 2.0, where relations can be represented as continuous or occurrent, with separate relation types for them (Smith et al., 2015). BFO's approach has been criticized for its complexity, that causes logic and usability issues (Mungall, 2013).

Acknowledgements

The development of Bio CRM was started in the EU COST project "Reassembling the Republic of Letters"¹⁶ (Tuominen et al., 2018).

6. References

- Aba Sah Dadzie and Matthew Rowe. 2011. Approaches to visualising Linked Data: A survey. *Semantic Web*, 2(2):89–124.
- Bruce D'Arcus and Frédérick Giasson. 2009. Bibliographic Ontology specification. <http://bibliontology.com>.
- Ian Davis. 2004. Relationship: A vocabulary for describing relationships between people. <http://vocab.org/relationship/>.
- Martin Doerr. 2009. Ontologies for cultural heritage. In S. Staab and R. Studer, editors, *Handbook on ontologies (2nd Edition)*, pages 463–486. Springer-Verlag.
- David Easley and Jon Kleinberg. 2010. *Networks, Crowds, and Markets: Reasoning about a Highly Connected World*. Cambridge University Press.
- Nicola Guarino and Christopher Welty. 2002. Evaluating ontological decisions with OntoClean. *Communications of the ACM*, 45(2):61–65.

¹⁶<http://www.republicofletters.net>

- Robert A. Hanneman and Mark Riddle. 2005. *Introduction to social network methods*. University of California, Riverside, CA.
- Frederik Hogenboom, Flavius Frasincar, Uzey Kaymak, and Franciska de Jong. 2011. An overview of event extraction from text. In *DeRiVE 2011, Detection, Representation, and Exploitation of Events in the Semantic Web at the 10th International Semantic Web Conference 2011 (ISWC 2011)*, Bonn, Germany, October.
- Eero Hyvönen, Olli Alm, and Heini Kuitinen. 2007. Using an ontology of historical events in semantic portals for cultural heritage. In *Proceedings of the Cultural Heritage on the Semantic Web Workshop at the 6th International Semantic Web Conference (ISWC 2007)*, Busan, Korea, November.
- Eero Hyvönen. 2012. *Publishing and using cultural heritage linked data on the semantic web*. Morgan & Claypool, Palo Alto, CA.
- Eero Hyvönen, Petri Leskinen, Erkki Heino, Jouni Tuominen, and Laura Sirola. 2017. Reassembling and enriching the life stories in printed biographical registers: Norssi high school alumni on the semantic web. In *Proceedings, Language, Technology and Knowledge (LDK 2017)*, pages 113–119, Galway, Ireland, June.
- Eero Hyvönen, Petri Leskinen, Minna Tamper, Jouni Tuominen, and Kirsi Keravuori. 2018. Semantic national biography of Finland. In *Proceedings of the Digital Humanities in the Nordic Countries 3rd Conference (DHN 2018)*, Helsinki, Finland, March.
- Matt Jensen. 2003. Visualising complex semantic timelines. NewsBlip Research Papers, Report NBTR2003-001. <http://www.newsblip.com/tr/>.
- Katharine S. B. Keats-Rohan. 2007. Biography, identity and names: Understanding the pursuit of the individual in prosopography. In *Prosopography Approaches and Applications. A Handbook*, pages 139–182. University of Oxford.
- Johannes Kehrner and Helwig Hauser. 2013. Visualization and visual analysis of multifaceted scientific data: A survey. *IEEE transactions on visualization and computer graphics*, 19(3):495–513.
- Kouji Kozaki, Eiichi Sunagawa, Yoshinobu Kitamura, and Riichiro Mizoguchi. 2006. Fundamental consideration of role concepts for ontology evaluation. In *Proceedings of EON2006*, Edinburgh, United Kingdom.
- Timothy Lebo, Satya Sahoo, and Deborah McGuinness. 2013. PROV-O: The PROV Ontology. W3C Recommendation 30 April 2013, <http://www.w3.org/TR/2013/REC-prov-o-20130430/>.
- Petri Leskinen, Jouni Tuominen, Erkki Heino, and Eero Hyvönen. 2017. An ontology and data infrastructure for publishing and using biographical linked data. In *Proceedings of the Workshop on Humanities in the Semantic Web (WHiSe II)*, Vienna, Austria, October.
- Petri Leskinen, Eero Hyvönen, and Jouni Tuominen. 2018. Analyzing and visualizing prosopographical linked data based on short biographies. In *Proceedings of Biographical Data in a Digital World 2017 (BD2017)*.
- Christopher J. Mungall. 2013. A critique of temporalized relations. Technical report. v1.3. April 25, <https://github.com/cmungall/trel-crit/raw/master/trc.pdf>.
- Gabor Nagypal, Richard Deswarte, and Jan Oosthoek. 2005. Applying the semantic web: The VICODI experience in creating visual contextualization for history. *Lit Linguist Computing*, 20(3):327–349.
- Yves Raimond and Samer Abdallah. 2007. The Event Ontology. <http://motools.sourceforge.net/event/event.html>.
- Dave Reynolds. 2014. The Organization Ontology. W3C Recommendation 16 January 2014, <https://www.w3.org/TR/2014/REC-vocab-org-20140116/>.
- Brian Roberts. 2002. *Biographical Research*. Understanding social research. Open University Press.
- Ansgar Scherp and Vasileios Mezaris. 2014. Survey on modeling and indexing events in multimedia. *Multimedia Tools and Applications*, 70(1):7–23, May.
- Ansgar Scherp, Thomas Franz, Carsten Saathoff, and Steffen Staab. 2009. F—a model of events based on the foundational ontology DOLCE+DnS Ultralight. In *Proceedings of the Fifth International Conference on Knowledge Capture (K-CAP '09)*, pages 137–144, Redondo Beach, California, USA.
- Ryan Shaw. 2010. LOD: An ontology for linking open descriptions of events. <http://linkedevents.org/ontology/>.
- Barry Smith, Mauricio Almeida, Jonathan Bona, Mathias Brochhausen, Werner Ceusters, Melanie Courtot, Randall Dipert, Albert Goldfain, Pierre Grenon, Janna Hastings, William Hogan, Leonard Jacuzzo, Ingvar Johansson, Chris Mungall, Darren Natale, Fabian Neuhaus, James Overton, Anthony Petosa, Robert Rovetto, Alan Ruttenberg, Mark Ressler, Ron Rudniki, Selja Seppälä, Stefan Schulz, and Jie Zheng. 2015. Basic Formal Ontology 2.0 – specification and user’s guide. Technical report. June 26.
- Jouni Tuominen, Eetu Mäkelä, Eero Hyvönen, Arno Bosse, Miranda Lewis, and Howard Hotson. 2018. Reassembling the republic of letters – a linked data approach. In *Proceedings of the Digital Humanities in the Nordic Countries 3rd Conference (DHN 2018)*, Helsinki, Finland, March.
- Willem Robert van Hage, Véronique Malaisé, Roxane Segers, Laura Hollink, and Guus Schreiber. 2011. Design and use of the simple event model (SEM). *Web Semantics: Science, Services and Agents on the World Wide Web*, 9(2):128–136.
- Koenraad Verboven, Myriam Carlier, and Jan Dumolyn. 2007. A short manual to the art of prosopography. In *Prosopography Approaches and Applications. A Handbook*, pages 35–70. University of Oxford.
- Jacopo Zingoni. 2014. PROSO data model – a solution for modelling historical academic prosopographical records as linked data through an event based ontological approach. In *Atelier Heloise Workshop*, Lyon, France, March.